(54) Title: METABOTROPIC GLUTAMATE RECEPTOR–LIKE PROTEIN AND ENCODING CDNA

(57) Abstract

The present invention provides an MGlur–like receptor protein (MGrcm) and polynucleotides which identify and encode MGrcm. The invention also provides expression vectors and host cells comprising nucleic acid sequence encoding MGrcm. The invention also provides antibodies of MGrcm and methods of diagnosing and treating diseases associated with expression of MGrcm, as well as screening assays employing the protein, nucleotide, and antibody compositions.
METABOTROPIC GLUTAMATE RECEPTOR-LIKE PROTEIN AND ENCODING CDNA

FIELD OF THE INVENTION

This invention relates to a novel MGlur-like receptor protein and polynucleotide compositions, to the production of these compositions, and to the use of the compositions in the diagnosis, prevention, and treatment of disease states.

BACKGROUND OF THE INVENTION

Receptors belonging to the 7-transmembrane G-protein coupled receptor (7-TDGR) superfamily are transmembrane proteins present in the plasma membrane characterized by amino acid sequences which contain seven hydrophobic domains predicted to form the membrane-spanning regions. 7-TDGRs transmit extracellular stimuli to the interior of the cell via interactions with G-proteins. The stimuli for different receptors in the superfamily include light, taste, odor, small peptides, amino acid derivatives and lipid analogs (see, e.g., Watson and Arkinstall (1994) The G-Protein Linked Receptor Factsbook, Academic Press, New York). Activation of the 7-TDGR receptor by extracellular stimuli leads to activation of intracellular G-proteins, which in turn modulates production of intracellular second messenger molecules such as cAMP and IP₃.
One class of receptors belonging to the 7-TDGR superfamily are the metabotropic glutamate receptors (mGluRs), which are involved in the stimulation of phospholipase C, the presynaptic inhibition of glutamate release, the closure of cation channels in retinal ON bipolar cells, and the modulation of adenylate cyclase. Within the mGluRs, eight subtypes have been identified, which share common structural architecture. The mGluRs contain a signal sequence (cleaved during post-translational processing), an extracellular NH₂-domain, seven transmembrane segments, and a cytosolic C-terminal domain. In some mGluR receptors, alternative splicing results in the production of receptor splice variants (Makoff A.J. et al., NeuroReport 8:2943-2947 (1997)).

The eight mGluR subtypes have been further classified into three groups based on sequence and functional similarities: Group I contains mGluR1 (and splice variants 1a, 1b, 1c, 1d, and 1e) and mGluR5 (and splice variants 5a and 5b); Group II contains mGluR2 and mGluR3; and Group III contains mGluR4 (and splice variants 4a and 4b), mGluR6, mGluR7 (and splice variants 7a and 7b) and mGluR8 (Nicoletti, F. et al., Trends Neurosci. 19:219-224 (1996)).

Consistent with their sequence identities, mGluRs within groups show similar (but not identical) biochemical properties. The group I receptors stimulate inositol triphosphate (IP₃) production and intracellular Ca²⁺ mobilization, induce arachidonic acid release, and have nearly identical agonist selectivities (Nakanishi, S., Science 258:597-603 (1992)). The group II and III receptors, on the other hand, inhibit the forskolin-stimulated accumulation of intracellular cAMP in an agonist-dependent manner, but with differing agonist selectivities between the two groups. The mGluR2, mGluR3, and mGluR4 receptors are all sensitive to PTX, indicating that the Group II and III receptors are coupled to inhibitory G (Gᵢ) proteins. Furthermore, the various mGluR subtypes show overlapping but non-equivalent expression patterns in mammalian brain. It is notable that receptor subtypes within the same group, though having similar primary sequences, signal transduction properties, and agonist
selectivities, are found to be differentially expressed in the CNS (Nakanishi, supra).

Glutaminergic neurotransmission is disrupted in many neuropathologic conditions. Imbalances in glutaminergic function have been implicated in neuronal death associated with ischemia, anoxia, stroke, epilepsy, and in neurodegenerative disorders. During brain ischemia and hypoglycemia, neuronal deterioration and cell death results from massive stimulation of glutamate receptors by high concentrations of extracellular glutamate. Glutamate neurotoxicity (or "excitotoxicity") may also underlie slow-progressing neurodegenerative diseases such as Huntington's disease, Alzheimer's disease, and Parkinson's disease.

Receptor proteins and the nucleic acids which encode them have various industrial applications, including as pharmaceutical and diagnostic agents. Receptor proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor-ligand interaction. The discovery of a novel receptor protein having sequence similarity to receptors of the mGluR family, and the polynucleotides which encode it, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of various disease states, such as disorders of the central and peripheral nervous systems, including neurological and neurodegenerative disorders, as well as cardiac, urologic, and gastrointestinal disorders.

SUMMARY OF THE INVENTION

The invention includes a novel, isolated receptor protein, having sequence similarity to the mGluR class of 7-transmembrane G-protein coupled receptors and identified herein as MGRcm, which comprises a sequence having at least 80 percent sequence identity with SEQ ID NO:2. In other embodiments, the protein comprises a sequence at having at least 90% identity to SEQ ID NO:2, or
comprises the sequence identified as SEQ ID NO:2. The invention also includes fragments of MGRcm, preferably at least about 10-50 amino acids in length, and as well as pharmaceutical compositions containing MGRcm.

By another embodiment the present invention concerns the mouse homolog of MGRcm and concerns a sequence having at least 80% identity, preferably 90% identity to the sequence identified as SEQ ID NO:4, fragments of said sequences having 10-50 amino acids and pharmaceutical compositions.

In another aspect the invention includes an isolated polynucleotide having a sequence which encodes MGRcm (both human and murine) as described above, or a sequence complementary to the MGRcm coding sequence, and a composition comprising the polynucleotide. The polynucleotide may be mRNA, DNA, cDNA, genomic DNA, as well as an antisense analog thereof. The polynucleotide may encode an MGRcm having at least 80% sequence identity to the protein sequence SEQ ID NO:2 or the sequence of SEQ ID NO:4. The polynucleotide may contain, for example, a coding sequence having at least 80% sequence identity with the polynucleotide sequence identified as SEQ ID NO:1 (coding for SEQ ID NO:2) or the nucleotide sequence identified as SEQ ID NO:3 (coding for SEQ. ID NO:4). In a specific embodiment, the polynucleotide has the sequence identified as SEQ ID NO:1. The composition also contemplates fragments of the polynucleotide, preferably at least about 15-30 nucleotides in length.

Also disclosed is a recombinant expression vector containing a polynucleotides as described above, and, operably linked to the polynucleotide, regulatory elements effective for expression of the protein in a selected host. Preferred coding sequences are given above. In a related aspect, the invention includes a host cell containing the vector.

The invention further includes a method for producing MGRcm by recombinant techniques, by culturing recombinant prokaryotic or eukaryotic host cells containing nucleic acid sequences encoding MGRcm under conditions
promoting expression of the protein, and subsequent recovery of the protein from the host cell or the cell culture medium.

In still another aspect, the invention includes an antibody specific against MGRcm. The antibody has diagnostic and therapeutic applications, particularly in treating neurologic and psychiatric disorders and neurodegenerative conditions, such as seizures, epilepsy, amnesia, dementia, migraine, depression, mania, Alzheimer's disease, Huntington's disease, Parkinson's disease, disorders associated with the peripheral nervous system, as well as cardiac, urologic, and gastrointestinal disorders.

Treatment methods which employ antisense or coding sequence polynucleotides for inhibiting or enhancing levels of MGRcm are also contemplated, as are treatment methods which employ antibodies specific against MGRcm.

Diagnostic methods for detecting levels of MGRcm in specific tissue samples, and for detecting levels of expression of the MGRcm in tissues, also form part of the invention. In one embodiment, a method of detecting a polynucleotide which encodes MGRcm in a biological sample, involves the steps of: (a) hybridizing the complement of a polynucleotide which encodes MGRcm to nucleic acid material of a biological sample, thereby forming a hybridization complex, and (b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding MGRcm in the biological sample. Methods for detecting mutations in the coding region of MGRcm are also contemplated.

Screening methods which employ MGRcm for identifying a candidate compound capable of binding to and modulating the activity of MGRcm also form part of the invention. An exemplary method includes (a) contacting a test compound with MGRcm, (b) measuring the effect of the test compound on the activity of the MGRcm, and (c) selecting the test compound as a candidate compound if its effect on the activity of MGRcm is above a selected threshold.
level. The activity measured may be, for example, MGRcm-mediated production of a second messenger such as cAMP or IP₃. In one embodiment, the test compound is a component of a combinatorial library. In another embodiment, the test compound is an antibody specific against MGRcm protein.

The invention also includes, in a related aspect, a compound identified by the screening methods described above.

These and other objects and features of the invention will become more fully apparent when the following detailed description is read in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1A shows the nucleic acid sequence (SEQ ID NO:1) and Fig. 1B shows the translated protein sequence (SEQ ID NO:2) of MGRcm;

Fig. 2A shows the nucleic acid sequence (SEQ ID NO:3) and Fig. 2B shows the translated protein sequence (SEQ ID NO:4) of a mouse homolog of MGRcm;

Fig. 3A shows a pairwise amino acid sequence alignment between MGRcm (SEQ ID NO:2) and human mGluR1 (MGR1_HUMAN, GenBank Accession Q13255);

Fig. 3B shows a pairwise amino acid sequence alignment between MGRcm (SEQ ID NO:2) and rat mGluR3 (MGR3_RAT, GenBank Accession P31422);

Fig. 4 shows a hydrophobicity analysis of SEQ ID NO:2, prepared using the TMpred program and the TMbase database.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO:1 is the nucleic acid sequence of the MGRcm transcript;
SEQ ID NO:2 is the predicted amino acid translation of MGRcm;
SEQ ID NO:3 is the nucleic acid sequence of the mouse homolog MGRcm transcript;
SEQ ID NO:4 is the predicted amino acid translation of transcript of the mouse homolog of MGRcm;
SEQ ID NO:5 is the amino acid sequence of human mGluR1 (MGR1_HUMAN, GenBank Accession Q13255);
SEQ ID NO:6 is the amino acid sequence of rat mGluR3 (MGR3_RAT, GenBank Accession P31422); and
SEQ ID NO:7 is an oligonucleotide used for northern blot analysis.

DETAILED DESCRIPTION OF THE INVENTION
I. Definitions

Unless otherwise indicated, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (Second Edition), Cold Spring Harbor Press, Plainview, N.Y. and Ausubel FM et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

An "mGluR-like" protein is a protein which has at least 20% amino acid sequence identity to a corresponding aligned region of at least one of the eight known mGluR subtypes.

"MGRcm" refers to an mGluR-like protein which contains a sequence having at least 80 percent, preferably 90 percent, and more preferably 95 percent sequence identity with the polypeptide identified as SEQ ID NO:2. The protein may be a mature MGRcm protein and/or a modified MGRcm protein. As used herein, reference to MGRcm is meant to include the full-length molecule and fragments thereof unless the context indicates otherwise.
The term also refers to the mouse homolog of MGRcm as depicted in SEQ ID NO:4, as well as to sequences having 80% identity, preferably 90% identity, most preferably 95% identity with this sequence.

The term "mature MGRcm protein" refers to the MGRcm receptor protein as it exists in the cell after post-translational processing, e.g. removal of a signal sequence.

The term "modified", when referring to a protein of the invention, means a protein which is modified either by natural processes, such as processing or other post-translational modifications, or by chemical modification techniques which are well known in the art. Among the numerous known modifications which may be present include, but are not limited to, acetylation, acylation, amidation, ADP-ribosylation, glycosylation, GPI anchor formation, covalent attachment of a lipid or lipid derivative, methylation, myristylation, pegylation, prenylation, phosphorylation, ubiquitination, or any similar process.

The term "biologically active" refers to an MGRcm having structural, regulatory or biochemical functions of the naturally occurring MGRcm, and the ability to modulate intracellular signal transduction, e.g., by the production of intracellular second messengers. Likewise, "immunologically active" defines the capability of a natural, recombinant or synthetic MGRcm, or any fragment thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "fragment", when referring to MGRcm, means a polypeptide which has an amino acid sequence which is the same as part of but not all of the amino acid sequence of an MGRcm receptor, which either retains essentially the same biological function or activity as MGRcm, or retains at least one of the functions or activities of MGRcm; for example, an extracellular fragment which retains the ability to bind an extracellular ligand of MGRcm, or an intracellular C-terminal fragment which retains signal transduction modulation activity, or a fragment which retains immunological activity of MGRcm. The fragment
preferably includes at least 10-50 contiguous residues of MGRcm, more preferably at least 50-200 residues, still more preferably at least 200-500 residues.

The term "portion", when referring to a protein of the invention, means a polypeptide which has an amino acid sequence which is the same as part of the amino acid sequence of the present invention or a variant thereof, which does not necessarily retain any biological function or activity.

A "conservative substitution" refers to the substitution of an amino acid in one class by an amino acid in the same class, where a class is defined by common physicochemical amino acid sidechain properties and high substitution frequencies in homologous proteins found in nature (as determined, e.g., by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). Six general classes of amino acid sidechains, categorized as described above, include: Class I (Cys); Class II (Ser, Thr, Pro, Ala, Gly); Class III (Asn, Asp, Gln, Glu); Class IV (His, Arg, Lys); Class V (Ile, Leu, Val, Met); and Class VI (Phe, Tyr, Trp). For example, substitution of an Asp for another class III residue such as Asn, Gln, or Glu, is a conservative substitution.

A "non-conservative substitution" refers to the substitution of an amino acid in one class with an amino acid from another class; for example, substitution of an Ala, a class II residue, with a class III residue such as Asp, Asn, Glu, or Gln.

"Optimal alignment" is defined as an alignment giving the highest percent identity score. Such alignment can be performed using a variety of commercially available sequence analysis programs, such as the local alignment program LALIGN using a kntup of 1, default parameters and the default PAM. A preferred alignment is the one performed using the CLUSTAL-W program from MacVector, operated with an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM similarity matrix. If a gap needs to be inserted into a first sequence to optimally align it with a second sequence, the percent identity is
calculated using only the residues that are paired with a corresponding amino acid residue (i.e., the calculation does not consider residues in the second sequences that are in the "gap" of the first sequence).

"Percent sequence identity", with respect to two amino acid or polynucleotide sequences, refers to the percentage of residues that are identical in the two sequences when the sequences are optimally aligned. Thus, 80% amino acid sequence identity means that 80% of the amino acids in two or more optimally aligned polypeptide sequences are identical.

A first polypeptide region is said to "correspond" to a second polypeptide region when the regions are essentially co-extensive when the sequences containing the regions are aligned using a sequence alignment program, as above. Corresponding polypeptide regions typically contain a similar, if not identical, number of residues. It will be understood, however, that corresponding regions may contain insertions or deletions of residues with respect to one another, as well as some differences in their sequences.

The term "sequence identity" means nucleic acid or amino acid sequence identity in two or more aligned sequences, aligned as defined above.

"Sequence similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. Thus, 80% protein sequence similarity means that 80% of the amino acid residues in two or more aligned protein sequences are conserved amino acid residues, i.e. are conservative substitutions.

The term "gene" as used herein means the segment of DNA involved in producing a polypeptide chain; it may include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).
An "isolated polynucleotide containing a sequence which encodes MGRcm" is a polynucleotide which contains the coding sequence of MGRcm (i) in isolation, (ii) in combination with additional coding sequences, such as fusion protein or signal peptide, in which the MGRcm coding sequence is the dominant coding sequence, (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host, and/or (iv) in a vector or host environment in which the MGRcm coding sequence is a heterologous gene.

The terms "heterologous DNA" and "heterologous RNA" refer to nucleotides that are not endogenous to the cell or part of the genome in which they are present; generally such nucleotides have been added to the cell, by transfection, microinjection, electroporation, or the like. Such nucleotides generally include at least one coding sequence, but this coding sequence need not be expressed.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The term "fragment", when referring to an MGRcm coding sequence, means a polynucleotide which has a nucleic acid sequence which is the same as part of but not all of the nucleic acid sequence of the MGRcm coding sequence. The polynucleotide fragment preferably includes at least 15 contiguous bases of MGRcm coding sequence, preferably at least 20-30 bases.
The term "expression vector" refers to vectors that have the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

The term "substantially purified" refers to molecules, either polynucleotides or polypeptides, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "variant" polynucleotide sequence encodes a "variant" amino acid sequence which is altered by one or more amino acids from the reference polypeptide sequence. The variant polynucleotide sequence may encode a variant amino acid sequence which contains "conservative" substitutions, wherein the substituted amino acid has structural or chemical properties similar to the amino acid which it replaces. In addition, or alternatively, the variant polynucleotide sequence may encode a variant amino acid sequence which contains "non-conservative" substitutions, wherein the substituted amino acid has dissimilar structural or chemical properties to the amino acid which it replaces. Variant polynucleotides may also encode variant amino acid sequences which contain amino acid insertions or deletions, or both.

An "allelic variant" is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

A "mutant" amino acid or polynucleotide sequence is a variant amino acid sequence, or a variant polynucleotide sequence which encodes a variant amino acid sequence, which has significantly altered biological activity from that of the naturally occurring protein.
A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

The term "modulate" as used herein refers to the change in activity of the polypeptide of the invention. Modulation may relate to an increase or a decrease in biological activity, binding characteristics, or any other biological, functional, or immunological property of the molecule.

The term "agonist" as used herein, refers to a molecule which, when bound to the receptor of the present invention, modulates the activity of the receptor by inducing, increasing, or prolonging the duration of the biological activity mediated by the receptor. Agonists may themselves be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of the receptor.

The term "antagonist" as used herein, refers to a molecule which, when bound to the receptor of the present invention, modulates the activity of the receptor by blocking, decreasing, or shortening the duration of the biological activity mediated by the receptor. Antagonists may themselves be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of the receptor.

The term "humanized antibody" refers to antibody molecule in which one or more amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding activity of the antibody.
"Treating a disease" refers to administering a therapeutic substance effective to reduce the symptoms of the disease and/or lessen the severity of the disease.

II. Polynucleotides Encoding MGRcm Proteins

The invention provides an isolated MGRcm protein and an isolated polynucleotide encoding the protein. As defined more fully in Section III below, MGRcm is (i) an mGluR-like protein, and (ii) contains an amino acid sequence having at least 80%, preferably 90% or 95%, sequence identity to the amino acid sequence identified as SEQ ID NO:2 or SEQ ID NO:4.

Fig. 1 shows a polynucleotide sequence encoding MGRcm, the polynucleotide identified herein as SEQ ID NO:1, and its translation product, SEQ ID NO:2. Fig. 2 shows a polynucleotide sequence encoding a mouse homolog of MGRcm, the polynucleotide identified herein as SEQ ID NO:3, and its translation product, SEQ ID NO:4.

Figs. 3A and 3B show pairwise Smith-Waterman amino acid sequence alignments between MGRcm (SEQ ID NO:2) and human mGluR1 (MGR1_HUMAN, GenBank Accession Q13255), and between MGRcm (SEQ ID NO:2) and rat mGluR3 (MGR3_RAT, GenBank Accession P31422), respectively. As seen in these figures, MGRcm shares approximately 20% amino acid sequence identity with portions of these mGlu receptor proteins.
A. **Mouse homolog transcripts**

Mouse homologs to the human MGRcm were identified which encode homologous protein in mice. One such homolog polynucleotide is identified herein as SEQ ID NO:3.

The invention contemplates mRNA transcripts which encode MGRcm or its mouse homolog, and DNA equivalents of the RNA, including cDNA and chemically synthesized DNA.

B. **Polynucleotide compositions**

The polynucleotides of the invention include sequences which encode MGRcm and MGRcm mouse homolog sequences complementary to the protein coding sequence, and novel fragments of the polynucleotide. The polynucleotides may be in the form of RNA or in the form of DNA, and include messenger RNA, synthetic RNA and DNA, cDNA, and genomic DNA. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (anti-sense, complementary) strand.

In a general embodiment, the polynucleotide has at least 70%, preferably at least 80% or 90% sequence identity with the sequence identified as SEQ ID NO:1. In another embodiment, the polynucleotide has at least 70%, preferably at least 80% or 90% sequence identity with the sequence identified as SEQ ID NO:3.

The polynucleotides may include the coding sequence of MGRcm or its mouse homolog (i) in isolation, (ii) in combination with additional coding sequences, such as fusion protein or signal peptide, in which the MGRcm coding sequence is the dominant coding sequence, (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the
coding sequence in a suitable host, and/or (iv) in a vector or host environment in which the MGRcm coding sequence is a heterologous gene.

The polynucleotide may encode a soluble fragment of MGRcm (human or murine), which includes the N-terminal extracellular fragment or the C-terminal intracellular fragment of the protein which have been cleaved from the transmembrane domains of the MGRcm receptor. The polynucleotides of the present invention may also have the protein coding sequence fused in-frame to a marker sequence which allows for purification of MGRcm or its mouse homolog. The marker sequence may be, for example, a hexahistidine tag to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell 37:767 (1984)).

Also contemplated are novel uses of polynucleotide fragments, also referred to herein as oligonucleotides, typically having at least 15 bases, preferably 20-30 bases corresponding to a region of the coding-sequence polynucleotide. The fragments may be used as probes, primers, antisense agents, and the like, according to known methods.

C. Preparation of polynucleotides

The polynucleotides may be obtained by screening cDNA libraries using oligonucleotide probes which can hybridize to or PCR-amplify polynucleotides which encode the MGRcm receptors and fragments disclosed above. cDNA libraries prepared from a variety of tissues are commercially available and procedures for screening and isolating cDNA clones are well-known to those of skill in the art. Such techniques are described in, for example, Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring

The polynucleotides may be extended to obtain upstream and downstream sequences such as promoters, regulatory elements, and 5' and 3' untranslated regions (UTRs). Extension of the available transcript sequence may be performed by numerous methods known to those of skill in the art, such as PCR or primer extension (Sambrook et al., supra), or by the RACE method using, for example, the Marathon RACE kit (Clontech, Cat. # K1802-1).

Alternatively, the technique of "restriction-site" PCR (Gobinda et al. PCR Methods Applic. 2:318-22, (1993)), which uses universal primers to retrieve flanking sequence adjacent a known locus, may be employed. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al., Nucleic Acids Res. 16:8186, (1988)). The primers may be designed using OLIGO(R) 4.06 Primer Analysis Software (1992); National Biosciences Inc, Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M et al. (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an
engineered double-stranded sequence into a flanking part of the DNA molecule before PCR.

Another method which may be used to retrieve flanking sequences is that of Parker, J.D., et al., *Nucleic Acids Res.* **19**:3055-60, (1991)). Additionally, one can use PCR, nested primers and PromoterFinder™ libraries to "walk in" genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

The polynucleotides and oligonucleotides of the invention can also be prepared by solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined to form continuous sequences up to several hundred bases.

**D. Applications of polynucleotides**

The polynucleotide coding sequences and novel oligonucleotides of the invention have a variety of uses in (1) synthesis of MGRcm, (2) diagnostics, (3) gene mapping, and (4) therapeutics.

**D1. Synthesis of MGRcm proteins and fragments**

In accordance with the present invention, polynucleotide sequences which encode MGRcm, its mouse homolog, fragments of the protein, fusion proteins, or functional equivalents thereof, collectively referred to herein as "MGRcm", may be used in recombinant DNA molecules that direct the expression of MGRcm in appropriate host cells. Due to the inherent degeneracy of the genetic code, other
nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express MGRcm.

As will be understood by those of skill in the art, it may be advantageous to produce MGRcm-encoding nucleotide sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray, E. et al., Nuc Acids Res. 17:477-508, (1989)) can be selected, for example, to increase the rate of MGRcm polypeptide expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The polynucleotide sequences of the present invention can be engineered in order to alter an MGRcm coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, alterations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are also described in Sambrook, et al., (supra).

The present invention also relates to host cells which are genetically engineered with vectors of the invention, and the production of proteins and polypeptides of the invention by recombinant techniques. Host cells are
genetically engineered (i.e., transduced, transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the MGRcm gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art.

The polynucleotides of the present invention may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host. The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and related sub-cloning procedures are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis. Examples of such promoters include: LTR or SV40 promoter, the E. coli lac or trp promoter, the phage lambda PL promoter, and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation, and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. In addition, the expression vectors preferably contain one or more selectable marker genes to provide a
phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as described above, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* and *Spodoptera* Sf9; mammalian cells such as CHO, COS, BHK, HEK 293 or Bowes melanoma; adenoviruses; plant cells, etc. It is understood that not all cell lines will be capable of functional coupling of the receptor to the cell's second messenger systems. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein. The invention is not limited by the host cells employed.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for MGRcm. For example, when large quantities of MGRcm are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as Bluescript(R) (Stratagene), in which the MGRcm coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster *J. Biol. Chem.* 264:5503-5509, (1989)); pET vectors (Novagen, Madison WI); and the like.

In the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al.* (supra) and Grant *et al.* *Methods in Enzymology* 153:516-544, (1987).
In cases where plant expression vectors are used, the expression of a sequence encoding MGRcm polypeptide may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson et al., *Nature* **310**:511-514, (1984)) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu et al. *EMBO J.* **6**:307-311, (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al., *EMBO J.* **3**:1671-1680, (1984); Broglie et al., *Science* **224**:838-843, (1984)); or heat shock promoters (Winter J and Sinibaldi RM, *Results. Probl. Cell Differ.* **17**:85-105, (1991)) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry LE (1992) in McGraw Hill Yearbook of Science and Technology, McGraw Hill, New York, N.Y., pp 191-196; or Weissbach and Weissbach (1988) *Methods for Plant Molecular Biology*, Academic Press, New York, N.Y., pp 421-463.

MGRcm may also be expressed in an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* Sf9 cells or in *Trichoplusia* larvae. The MGRcm coding sequence is cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of MGRcm coding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which MGRcm is expressed (Smith et al., *J. Virol.* **46**:584 (1983); Engelhard EK et al., *Proc. Natl. Acad. Sci.* **91**:3224-7, (1994)).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, an MGRcm coding sequence may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite
leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing MGRcm in infected host cells (Logan and Shenk, Proc. Natl. Acad. Sci. 81:3655-59, (1984)). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of an MGRcm coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where MGRcm coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D et al., Results Probl. Cell Differ. 20:125-62, (1994); Bittner et al., Methods in Enzymol., 153:516-544, (1987)).

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., and Battey, I., Basic Methods in Molecular Biology, (1986)). Cell-free translation systems can also be employed to produce polypeptides using RNAs derived from the DNA constructs of the present invention.
A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, BHK, MDCK, 293, WI38, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express MGRcm may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M et al., Cell 11:223-32, (1977)) and adenine phosphoribosyltransferase (Lowy I et al., Cell 22:817-23, (1980)) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M et al. Proc. Natl. Acad. Sci. 77:3567-70, (1980)); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F. et al., J. Mol. Biol.
150:1-14, (1981)) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan, Proc. Natl. Acad. Sci. 85:8047-51, (1988)). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrates, luciferin and ATP, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA, et. Al. Methods. Mol. Biol. 55:121-131, (1995)).

Host cells transformed with a nucleotide sequence encoding MGRcm may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding MGRcm can be designed with signal sequences which direct secretion of MGRcm polypeptide through a prokaryotic or eukaryotic cell membrane.

MGRcm may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle, Wash.). The inclusion of a protease-cleavable polypeptide linker sequence between the purification domain and MGRcm is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising MGRcm (e.g., a
soluble MGRcm fragment) fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography, as described in Porath et al., Protein Expression and Purification 3:263-281, (1992)) while the enterokinase cleavage site provides a means for isolating MGRcm from the fusion protein. pGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well known to those skilled in the art.

MGRcm can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.
D2. Diagnostic applications

The polynucleotides of the present invention may be used for a variety of diagnostic purposes. The polynucleotides may be used to detect and quantitate expression of MGRcm or its splice variants in patient's cells, e.g. biopsied tissues, by detecting the presence of mRNA coding for MGRcm receptor or its splice variants. This assay typically involves obtaining total mRNA from the tissue and contacting the mRNA with a nucleic acid probe. The probe is a nucleic acid molecule of at least 15 nucleotides, preferably 20-30 nucleotides, capable of specifically hybridizing with a sequence included within the sequence of a nucleic acid molecule encoding MGRcm under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of MGRcm. This assay can be used to distinguish between absence, presence, and excess expression of MGRcm and to monitor levels of MGRcm expression during therapeutic intervention. This assay can also be used to identify the presence of alternatively spliced transcripts of MGRcm, by designing multiple probes corresponding to different exons, and/or probes which preferentially hybridize to particular exon/intron interfaces.

The invention also contemplates the use of the polynucleotides as a diagnostic for diseases resulting from inherited defective MGRcm genes. These genes can be detected by comparing the sequences of the defective (i.e., mutant) MGRcm gene with that of a normal one. Association of a mutant MGRcm gene with abnormal MGRcm activity may be verified. In addition, mutant MGRcm genes can be inserted into a suitable vector for expression in a functional assay system (e.g., colorimetric assay, or complementation experiments in an MGRcm-deficient strain of mammalian cells, e.g. non-neuronal cells such as CHO or BHK) as yet another means to verify or identify mutations. Once mutant genes have been identified, one can then screen populations of interest for carriers of the mutant gene.
Individuals carrying mutations in the gene of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's cells, including but not limited to such as from blood, urine, saliva, placenta, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, et al., Nature 324:163-166, (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the gene of the present invention. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype.

Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such RNase and S1 protection or the chemical cleavage method (e.g. Cotton, et al., Proc. Natl. Acad. Sci. USA 85:4397-4401, (1985)), or by differences in melting temperatures. "Molecular beacons" (Kostrikis L.G. et al., Science 279:1228-1229, (1998)), hairpin-shaped, single-stranded synthetic oligonucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of MGRcm. Such diagnostics would be particularly useful for prenatal testing.

Another method for detecting mutations uses two DNA probes which are designed to hybridize to adjacent regions of a target, with abutting bases, where the region of known or suspected mutation(s) is at or near the abutting bases. The two probes may be joined at the abutting bases, e.g., in the presence of a ligase enzyme, but only if both probes are correctly base paired in the region of
probe junction. The presence or absence of mutations is then detectable by the presence or absence of ligated probe.

Also suitable for detecting mutations in the MGRcm coding sequence are oligonucleotide array methods based on sequencing by hybridization (SBH), as described, for example, in U.S. Patent No. 5,547,839. In a typical method, the DNA target analyte is hybridized with an array of oligonucleotides formed on a microchip. The sequence of the target can then be "read" from the pattern of target binding to the array.

D3. Gene mapping

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the MGRcm cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, which would complicate the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an
analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include in situ hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

5 Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma et al. (1988) *Human Chromosomes: a Manual of Basic Techniques*, Pergamon Press, New York.

10 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in the OMIM database (Center for Medical Genetics, Johns Hopkins University, Baltimore, MD and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD). The OMIM gene map presents the cytogenetic map location of disease genes and other expressed genes. The OMIM database provides information on diseases associated with the chromosomal location. Such associations include the results of linkage analysis mapped to this interval, and the correlation of translocations and other chromosomal aberrations in this area with the advent of polygenic diseases, such as cancer.

D4. Therapeutic applications

Polynucleotides which encode MGRCm, or complements of the polynucleotides, may also be used for therapeutic purposes. Expression of MGRCm may be modulated through antisense technology, which controls gene expression through complementary polynucleotides, i.e. antisense DNA or RNA, to the control, 5' or regulatory regions of the gene encoding MGRCm. For example, the 5' coding portion of the polynucleotide sequence which codes for the protein of the present invention is used to design an antisense oligonucleotide of from about 10 to 40 base pairs in length. Oligonucleotides derived from the
transcription start site, e.g. between positions -10 and +10 from the start site, are preferred. An antisense DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (Lee et al., Nucl. Acids Res. 6:3073, (1979); Cooney et al., Science 241:456, (1988); and Dervan et al., Science 251:1360, (1991)), thereby preventing transcription and the production of MGRcm. An antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into MGRcm protein (Okano, J. Neurochem., 56:560, (1991)). The antisense constructs can be delivered to cells by procedures known in the art such that the antisense RNA or DNA may be expressed *in vivo*.

The therapeutic polynucleotides of the invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The polypeptides, and agonist and antagonist compounds which are polypeptides, may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy." Cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo*
and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors mentioned above may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317, psi-2, psi-AM, PA12, T19-14X, VT-19-17-H2, psi-CRE, psi-CRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, *Human Gene Therapy*, Vol. 1:5-14, (1990). The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and CaPO₄ precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either in vitro or in vivo. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic
carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

The genes introduced into cells may be placed under the control of inducible promoters, such as the radiation-inducible Egr-1 promoter, (Maceri, H.J., et al., Cancer Res, 56(19):4311 (1996)), to stimulate MGRcm production or antisense inhibition in response to radiation, e.g., radiation therapy for treating tumors.

III. MGRcm Receptor

Hydrophobicity analysis (Fig. 4) shows that the 565-residue MGRcm sequence (SEQ ID NO:2) contains up to 8 potential membrane-spanning domains in the N-terminal region of the protein. MGRcm possesses the characteristic architecture of a 7-transmembrane (7tm) receptor, with a putative signal peptide, a relatively short extracellular domain extending from about residue 20 to about residue 40, a group of 7 transmembrane segments extending from about residue 40 to about residue 300, and a relatively long C-terminal cytosolic domain extending from about residue 300 to residue 565.

Figures 3A and 3B show that MGRcm shares approximately 20% amino acid sequence identity with portions of human mGluR1 (MGR1_HUMAN, GenBank Accession Q13255) and rat mGluR3 (MGR3_RAT, GenBank Accession P31422). Furthermore, using the program Profilesearch, the region corresponding to residues 92-262 of SEQ ID NO:2 attained a highly significant normalized score (N score) of 10.7651 against the profile PS50259 (G_PROTEIN_RECEPTOR_F3_4; PROSITE database). This score is above the N score cutoff of 8.5 for this profile, providing further evidence that MGRcm is a novel member of the metabotropic glutamate / GABA receptor class of 7tm G-protein coupled receptors.

Northern blots (Cat. No. HB-1020, OriGene Technologies, Inc., Rockville, Maryland) were probed using the polynucleotide SEQ ID NO:7 according to the
protocol supplied by the manufacturer. The northern analysis detected the presence of MGRcm coding sequence in brain, liver, kidney, stomach, and heart.

The substantially purified MGRcm of the invention includes a protein containing an amino acid sequence having at least 80%, preferably at least 90% or 95% identity to the sequence identified as SEQ ID NO:2 or SEQ ID NO:4. The protein may be a recombinant protein, a natural protein or a synthetic protein, preferably a recombinant protein. The protein may be in mature and/or modified form, also as defined above. Also contemplated are protein fragments having at least 10-50 contiguous amino acid residues derived from MGRcm.

The sequence variations are preferably those that are considered conserved substitutions, as defined above. Thus, for example, a protein with a sequence having at least 80% sequence identity with the protein identified as SEQ ID NO:2 (565 amino acid residues) contains up to 113 amino acid substitutions, preferably conserved substitutions as defined above. In a more specific embodiment, the protein has or contains the sequence identified as SEQ ID NO:2 or SEQ ID NO:4. MGRcm may be (i) a protein in which one or more of the amino acid residues in a sequence listed above are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue), or (ii) a protein in which one or more of the amino acid residues includes a substituent group, or (iii) a protein in which the MGRcm is fused with another compound, such as a compound to increase the half-life of the protein (for example, polyethylene glycol (PEG)), or (iv) a protein in which additional amino acids are fused to MGRcm, or (v) an isolated fragment of the protein which is soluble, i.e. not membrane bound, yet still binds its natural ligands.

Such fragments, variants and derivatives are deemed to be within the scope of those skilled in the art from the teachings herein.
A. Preparation of MGRcm

Recombinant methods for producing and isolating MGRcm proteins, splice variants, and fragments are described above.

In addition to recombinant production, fragments and portions of MGRcm may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart et al. (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield, J., J. Am. Chem. Soc., 85:2149-2154, (1963)). In vitro peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer. Portions of MGRcm may be chemically synthesized separately and combined using chemical methods.

The receptor protein may also be obtained by isolation from natural sources, e.g., by affinity purification using the anti-MGRcm antibody described in the section below (either from humans or from mice cells).

B. Applications of MGRcm

The MGRcm receptor of the invention has uses in (1) therapeutic treatment methods and (2) drug screening.

B1. Therapeutic uses and compositions

Based on its similarity with the class of mGluR receptors, and its tissue distribution, MGRcm proteins of the present invention are generally useful in treating diseases and disorders which respond to the modulation of signal transduction via the central and/or peripheral nervous system. These include cardiac, urologic, and gastrointestinal disorders, neurodegenerative disorders, and neuronal degradation associated with, for example, ischemia, anoxia, stroke, and epilepsy, and other disorders relating to overstimulation of the sympathetic nervous system including hypoglycemia, vasoconstriction, renal failure,
arrhythmia, peripheral vascular disorders, heart failure, nephrosis, cirrhosis, dysphagia, and gastritis.

A polypeptide fragment of MGRcm, preferably a soluble fragment, may be employed to inhibit activity of MGRcm by binding an agonist which is essential for MGRcm activity and thus preventing the agonist from interacting with MGRcm. A fragment of MGRcm, preferably a soluble fragment, may alternatively be employed to block the binding of agonists to MGRcm, thus likewise preventing the cellular response induced by the binding of the agonist to the MGRcm receptor.

MGRcm compositions are tested in appropriate in vitro and in vivo animal models of disease, to confirm efficacy, tissue metabolism, and to estimate dosages, according to methods well known in the art.

MGRcm compositions may be administered by any of a number of routes and methods designed to provide a consistent and predictable concentration of compound at the target organ or tissue. The polypeptide compositions may be administered alone or in combination with other agents, such as stabilizing compounds, and/or in combination with other pharmaceutical agents such as drugs or hormones.

MGRcm compositions may be administered by a number of routes including, but not limited to oral, intravenous, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means. MGRcm compositions may also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

The polypeptide can be given via intravenous or intraperitoneal injection. Similarly, the polypeptide may be injected to other localized regions of the body. The polypeptide may also be administered via nasal insufflation. Enteral administration is also possible. For such administration, the polypeptide should be formulated into an appropriate capsule or elixir for oral administration, or into a suppository for rectal administration.
The foregoing exemplary administration modes will likely require that the polypeptides be formulated into an appropriate carrier, including ointments, gels, suppositories. Appropriate formulations are well known to persons skilled in the art.

Dosage of the polypeptide will vary, depending upon the potency and therapeutic index of the particular polypeptide selected. These parameters are easily determinable by the skilled practitioner. As an example, if the polypeptide inhibits neuronal cell degradation *in vitro* at a given concentration, the practitioner will know that the final desired therapeutic concentration will be this range, calculated on the basis of expected biodistribution. An appropriate target concentration is in the ng/kg to low mg/kg range, e.g., 50 ng/kg to 1 mg/kg body weight, for IV administration.

A therapeutic composition for use in the treatment method can include the polypeptide in a sterile injectable solution, the polypeptide in an oral delivery vehicle, or the polypeptide in a nebulized form, all prepared according to well known methods. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof.

B2. Screening methods

The present invention also includes an assay for identifying molecules, such as synthetic drugs, antibodies, peptides, or other molecules, which have a modulating effect on the activity of MGRcm, *e.g.* agonists or antagonists of the MGRcm receptor of the present invention. Such an assay comprises the steps of providing a functional MGRcm receptor encoded by the polynucleotides of the present invention, contacting the MGRcm receptor with one or more molecules to determine its modulating effect on the activity of the receptor, and selecting from the molecules a candidate molecule capable of modulating MGRcm
receptor activity. Such compounds are useful in the treatment of disease
conditions associated with activation or depression of MGRcm activity.

MGRcm, its catalytic or immunogenic fragments or oligopeptides thereof,
can be used for screening therapeutic compounds in any of a variety of drug
screening techniques. The protein employed in such a test may be
membrane-bound, free in solution, affixed to a solid support, borne on a cell
surface, or located intracellularly. The formation of binding complexes between
MGRcm and the agent being tested may be measured. Compounds which inhibit
binding between MGRcm and its agonists may also be measured.

In one embodiment, the screening system includes recombinantly
expressed MGRcm, and the compounds screened are tested for their ability to
block or enhance the signal transduction activity of MGRcm. In one example of
a functional screening assay, mammalian cell lines which lack MGRcm receptor
are used to express MGRcm which interacts with G protein(s) present in the cell
line. In this assay, compounds are screened for their relative affinity as receptor
agonists or antagonists by comparing the relative receptor occupancy to the
extent of ligand-induced stimulation or inhibition of second messenger
production, such as cAMP or IP3. Kits for quantitating the production of second
messenger molecules are commercially available (e.g., Amersham)

Another technique for drug screening which may be used provides for
high throughput screening of compounds having suitable binding affinity to the
MGRcm protein is described in detail by Geysen in PCT Application WO
84/03564, published on Sep. 13, 1984. In summary, large numbers of different
small peptide test compounds are synthesized on a solid substrate, such as plastic
pins or some other surface. The peptide test compounds are reacted with soluble
fragments of MGRcm, or intact MGRcm solubilized in detergents or in lipid
vesicles, and washed. Bound MGRcm is then detected by methods well known in
the art. Substantially purified MGRcm can also be coated directly onto plates for
use in the aforementioned drug screening techniques. Alternatively,
non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

Antibodies to the MGRcm, as described in Section IV. below, may also be used in screening assays according to methods well known in the art. For example, a "sandwich" assay may be performed, in which an anti-MGRcm antibody is affixed to a solid surface such as a microtiter plate and solubilized MGRcm is added. Such an assay can be used to capture compounds which bind to MGRcm. Alternatively, such an assay may be used to measure the ability of compounds to interfere with the binding of an MGRcm agonist to MGRcm.

IV. Anti-MGRcm antibodies

In still another aspect of the invention, purified MGRcm is used to produce anti-MGRcm antibodies which have diagnostic and therapeutic uses related to the activity, distribution, and expression of MGRcm.

Antibodies to MGRcm may be generated by methods well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments and fragments produced by an Fab expression library. Antibodies, i.e., those which block ligand binding, are especially preferred for therapeutic use.

MGRcm for antibody induction does not require biological activity; however, the protein fragment or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of a MGRcm polypeptide may be fused with another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to MGRcm.
For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with MGRcm protein or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal antibodies to MGRcm may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (1975; Nature 256:495-497), the human B-cell hybridoma technique (Kosbor et al., Immunol. Today 4:72, (1983); Cote et al., Proc. Natl. Acad. Sci., 80:2026-2030, (1983)) and the EBV-hybridoma technique (Cole, et al., Mol. Cell Biol. 62:109-120, (1984)).

Techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can also be used (Morrison et al., Proc. Natl. Acad. Sci. 81:6851-6855, (1984); Neuberger et al. Nature, 312:604-608, (1984); Takeda et al., Nature, 314:452-454, (1985)). Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single-chain antibodies specific for MGRcm.

Antibody fragments which contain specific binding sites for MGRcm may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse W.D., et al., Science 256:1275-1281, (1989)).

A. Diagnostic applications

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between MGRcm and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on MGRcm is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox D.E. et al., J. Exp. Med. 158:1211, (1983). Antibodies which specifically bind MGRcm protein are useful for the diagnosis of conditions or diseases characterized by expression of MGRcm. Alternatively, such antibodies may be used in assays to monitor patients being treated with MGRcm, its agonists, or its antagonists. Diagnostic assays for MGRcm protein include methods utilizing the antibody and a label to detect MGRcm in extracts of cells or tissues. The proteins and antibodies of the present invention may be used with or without modification. Frequently, the proteins and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known in the art.

A variety of protocols for measuring MGRcm, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA),
radioimmunoassay (RIA), and fluorescent activated cell sorting (FACS). As noted above, a two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on MGRcm is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, et al. (supra). Such protocols provide a basis for diagnosing altered or abnormal levels of MGRcm expression. Normal or standard values for MGRcm expression are established by combining cell extracts taken from normal subjects, preferably human, with antibody to MGRcm under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified by various methods, preferably by photometric methods. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

The antibody assays are useful to determine the level of MGRcm present in a particular tissue, e.g., biopsied tumor tissue or neuronal tissue, as an indication of whether MGRcm is being overexpressed or underexpressed in the tissue, or as an indication of how MGRcm levels are responding to drug treatment.

B. Therapeutic uses

Based on the structural similarity between the mGluR-like protein MGRcm and members of the mGluR receptor family, therapeutic value may be achieved by administering an antibody specific against MGRcm to inhibit the action of MGRcm by, e.g., inhibiting the binding of agonists to the MGRcm receptor, to treat conditions generally associated with overstimulation of the mGluR class of receptors. Such conditions include those generally classed as cognitive disorders, including dementia, delirium, and amnesic disorders. Other conditions amenable to MGRcm antibody therapy include neurodegenerative
diseases such as Huntington's disease, Alzheimer's disease, and Parkinson's disease; neuronal cell death associated with, for example, ischemia, anoxia, stroke, and epilepsy; and other disorders relating to overstimulation of the sympathetic nervous system including hypoglycemia, vasoconstriction, renal failure, arrhythmia, peripheral vascular disorders, heart failure, nephrosis, cirrhosis, dysphagia, and gastritis.

The antibody employed is preferably a humanized monoclonal antibody, or a human Mab produced by known globulin-gene library methods. The antibody is administered typically as a sterile solution by IV injection, although other parenteral routes may be suitable. Typically, the antibody is administered in an amount between about 1-15 mg/kg body weight of the subject. Treatment is continued, e.g., with dosing every 1-7 days, until a therapeutic improvement is seen.

In one exemplary method, the antibody treatment is used to treat neuronal cell death associated with sudden-onset conditions such as ischemia, anoxia, stroke, or epilepsy, by administering the antibody with the first observable signs of the episode. The antibody may also be used to treat epileptics or those prone to stroke, seizures, or hypoglycemia, by administering the antibody prior to the episodes. Treatment for chronic disorders such as long-term neurodegenerative conditions is also contemplated, in this case, with long term injection of antibody.

Although the invention has been described with reference to specific methods and embodiments, it is appreciated that various modifications and changes may be made without departing from the invention.
SEQUENCE LISTING

5 <110> COMPUGEN LTD.

10 <120> NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES

15 <130> compugen

<140>
<141>

20 <160> 7

25 <170> PatentIn Ver. 2.0

30 <210> 1

35 <211> 2406

<212> DNA

<213> Human adenovirus type 1

<400> 1

45 ngatgcngnc ggcaggttcc gagttgggca gaacctggaa cggtcacagg aaaccggggc 60
ggaggttagt cgggcoagagg gggacctttg cccacggttc agagtgttct ggactcggcc 120
cacgcgttag aggagcgttt ggtgacgatc ccacgtaca gc tttgtgacgt gtcggaggttc 180
cagagaagga cctggtcggtgct aaaaaagccgaa ccctcgggtgcct cggtgtgttc acgtgtgcaa 240
gctggggagg ggaggggggggc cagagggcgt gacaagacg gactctgttct tagtgtgtgct 300
cctcgcctcc gcccgcctccc agcgaacctgt gcctgaaaag ctgactoacg gagggccagg 360
acgaaggtga caaggttag tggttccccca cggagacgcg ccaaggtgcg cccgcgcgtcg 420
tccgtagggc cgcctctctgg aagacgcggt ggggggtgcg cagggctgca ccotcacacc 480
aattgcccccg gccaggttcag gccgaggttcag tctgctgccagc ggcgagttcagc 540
gtcggtggcc gcggcttcag ggtgtggttcag gcggcttcagc gcggcttcagc 600
ctttcctcac cagccggaaa gtaagagttcag gcggaggttcag gcggcttcagc 720
ttcctttcctct cagccggaaa gtaagagttcag gcggaggttcag gcggcttcagc 840
tcctctttcac cagccggaaa gtaagagttcag gcggaggttcag gcggcttcagc 900
ccttctttcct acacacagc cccacccacc cccacccacc cccacccacc 1020
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1140
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1200
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1260
ttttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1320
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1380
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1440
ttttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1500
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1560
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1620
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1680
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1740
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1800
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1860
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1920
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 1980
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 2040
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 2100
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 2160
ctttctttcct ctacacagc cccacccacc cccacccacc cccacccacc 2220
ctgccagag caccctgccc gtcacaaccc aagcaaatag tgttcctggg gttggtggctg 2280
gcgagcgcct atgatctctc ggagattct gcaacctcaa ggaacttccc aggcgctcaag 2340
gcgtggatct tgtgctctctg tgaggaacaa ggggctctaa taaatcatt tctgctttat 2400
taaaaa 2406

<210> 2
<211> 479
<212> PRT
<213> Human adenovirus type 1

<400> 2
Met Arg Gly Arg Gly Ser Gln Gln Gln Gln Pro Thr Arg Arg Gln Gly

Gln Lys Leu Pro Ser Ser Ser Pro Ala Gly Lys Tyr Glu Ser Ala Gln

Pro Gly Gly Thr Gln Pro Glu Pro Gly Leu Gly Ala Arg Met Ala Ile

His Lys Ala Leu Val Met Cys Leu Gly Leu Pro Leu Phe Leu Phe Pro

Gly Ala Trp Ala Gln Gly His Val Pro Pro Gly Cys Ser Gln Gly Leu

Asn Pro Leu Tyr Tyr Asn Leu Cys Asp Arg Ser Gly Ala Trp Gly Ile
Val Leu Glu Ala Val Ala Gly Ala Gly Ile Val Thr Thr Phe Val Leu
   100        105        110

Thr Ile Ile Leu Val Ala Ser Leu Pro Phe Val Gln Asp Thr Lys Lys
   115        120        125

Arg Ser Leu Leu Gly Thr Gln Val Phe Phe Leu Leu Gly Thr Leu Gly
   130        135        140

Leu Phe Cys Leu Val Phe Ala Cys Val Val Lys Pro Asp Phe Ser Thr
   145        150        155        160

Cys Ala Ser Arg Arg Phe Leu Phe Gly Val Leu Phe Ala Ile Cys Phe
   165        170        175

Ser Cys Leu Ala Ala His Val Phe Ala Leu Asn Phe Leu Ala Arg Lys
   180        185        190

Asn His Gly Pro Arg Gly Trp Val Ile Phe Thr Val Ala Leu Leu Leu
   195        200        205

Thr Leu Val Glu Val Ile Ile Asn Thr Glu Trp Leu Ile Ile Thr Leu
   210        215        220

Val Arg Gly Ser Gly Glu Gly Gly Gly Pro Gln Gly Asn Ser Ser Ala Gly
   225        230        235        240

Trp Ala Val Ala Ser Pro Cys Ala Ile Ala Asn Met Asp Phe Val Met
   245        250        255
Ala Leu Ile Tyr Val Met Leu Leu Leu Leu Gly Ala Phe Leu Gly Ala
5     260     265     270

Trp Pro Ala Leu Cys Gly Arg Tyr Lys Arg Trp Arg Lys His Gly Val
10    275     280     285

Phe Val Leu Leu Thr Thr Ala Thr Ser Val Ala Ile Trp Val Val Trp
15    290     295     300

Ile Val Met Tyr Thr Tyr Gly Asn Lys Gln His Asn Ser Pro Thr Trp
20    305     310     315     320

Asp Asp Pro Thr Leu Ala Ile Ala Leu Ala Ala Asn Ala Trp Ala Phe
25    325     330     335

Val Leu Phe Tyr Val Ile Pro Glu Val Ser Gln Val Thr Lys Ser Ser
30    340     345     350

Pro Glu Gln Ser Tyr Gln Gly Asp Met Tyr Pro Thr Arg Gly Val Gly
35    355     360     365

Tyr Glu Thr Ile Leu Lys Glu Gln Lys Gly Gln Ser Met Phe Val Glu
40    370     375     380

Asn Lys Ala Phe Ser Met Asp Glu Pro Val Ala Ala Lys Arg Pro Val
45    385     390     395     400

Ser Pro Tyr Ser Gly Tyr Asn Gly Gin Leu Leu Thr Ser Val Tyr Gln
50
Pro Thr Glu Met Ala Leu Met His Lys Val Pro Ser Glu Gly Ala Tyr

Asp Ile Ile Leu Pro Arg Ala Thr Ala Asn Ser Gln Val Met Gly Ser

Ala Asn Ser Thr Leu Arg Ala Glu Asp Met Tyr Ser Ala Gln Ser His

Gln Ala Ala Thr Pro Arg Glu Arg Arg Gln Glu Leu Ser Gly Leu

<210> 3
<211> 891
<212> DNA
<213> Human adenovirus type 1

<400> 3
aattcggatc catgccctgc tccagcccgg agcagcgaaac tgtgccaggg acccagctga 60
agcctggcct gggagccagg atggccacct acagaacctt gctgtatgtgc cyaggactgc 120
cctctttctt cccaggagcc ttggccccaga atcatgcccc acctggttgc agcgcagacg 180
tggatccccc ctactacaac ctcttgtgac gctcgggggg cttggggcatt gtcttggagg 240
cagtggctgg agcaggcact atcagcagat tgtgttaacc catcatcctt gttgctagco 300
ttccattgtg gcaggacact aagaagcgga gcctctttggg gacccaggtg tcttctcctg 360
tgggcacccg gggtctcttc tgcctcgtgt ttgcctgtgtg ggtgaaagcc gactcttctta 420
cctgtgcctc tgcagccttc tcttwgggg ttcgttttgc catc wgcttc cttctgtcgg 480
tagctcacgt cctttccttc aacctttctaa cccggagaagaa ccaggggacct cggggcgtgg 540
tgatctttcac ctggtgccgctg cttgctctaccc tgtgggaggt catcattaacc acggagttgcc 600
ttatcatcac ccttgtacctgc ggaggtggcc aggttagccc cctggcgaat gtcagtgccg 660
actcgacact gccctctccc gttgccatcg ccaacatgga ctttgctcatg gctctcactc 720
acgtaatggt gttgctgctg acggctcttc tagggacctg gcccacccctt tggggtgctg 780
tcaagcgctg cgccaacacac ggggtcctttt tgggctgctac cactgctcact cctcatgcaca 840
tctgggttggt atggatgtgc atgtacactt acggcaacga gcagcaccat a 891

<210> 4
<211> 296
<212> PRT
<213> Human adenovirus type 1

<400> 4

Phe Gly Ser Met Pro Arg Ser Ser Pro Glu Gln Arg Thr Val Pro Gly
1 5 10 15

Thr Gln Leu Lys Pro Gly Leu Gly Ala Arg Met Ala Thr His Arg Thr
20 25 30

Leu Leu Met Cys Xaa Gly Leu Pro Leu Phe Phe Pro Gly Ala Leu Ala
35 40 45

Gln Asn His Ala Pro Pro Gly Cys Ser Pro Asp Leu Asp Pro Leu Tyr
50 55 60

Tyr Asn Leu Cys Asp Arg Ser Gly Ala Trp Gly Ile Val Leu Glu Ala
55 65 70 75 80
Val Ala Gly Ala Gly Ile Ile Thr Thr Phe Val Leu Thr Ile Ile Leu
85 90 95

Val Ala Ser Leu Pro Phe Val Gln Asp Thr Lys Lys Arg Ser Leu Leu
100 105 110

Gly Thr Gln Val Phe Phe Leu Leu Gly Thr Leu Gly Leu Phe Cys Leu
115 120 125

Val Phe Ala Cys Val Val Lys Pro Asp Phe Ser Thr Cys Ala Ser Arg
130 135 140

Arg Phe Leu Xaa Gly Val Leu Phe Ala Ile Xaa Phe Ser Cys Leu Val
145 150 155 160

Ala His Val Leu Ser Leu Asn Phe Leu Thr Arg Lys Asn His Gly Pro
165 170 175

Arg Gly Trp Val Ile Phe Thr Val Ala Leu Leu Leu Thr Leu Val Glu
180 185 190

Val Ile Ile Asn Thr Glu Trp Leu Ile Ile Thr Leu Val Pro Gly Gly
195 200 205

Gly Gln Val Ser Pro Leu Ala Asn Val Ser Ala Asp Ser Thr Met Thr
210 215 220

Ser Pro Cys Ala Ile Ala Asn Met Asp Phe Val Met Ala Leu Ile Tyr
225 230 235 240
Val Met Leu Leu Leu Leu Thr Ala Phe Leu Gly Ala Trp Pro Thr Leu
5 245 250 255

Cys Cys Arg Phe Lys Arg Trp Arg Lys His Gly Val Phe Val Leu Leu
10 260 265 270

Thr Thr Val Ile Ser Ile Ala Ile Trp Val Val Trp Ile Val Met Tyr
15 275 280 285

Thr Tyr Gly Asn Glu Gln His His
20 290 295

<210> 5
<211> 1194
<212> PRT
<213> Human adenovirus type 1

40

<400> 5
Met Val Gly Leu Leu Leu Phe Phe Phe Pro Ala Ile Phe Leu Glu Val
1 5 10 15

Ser Leu Leu Pro Arg Ser Pro Gly Arg Lys Val Leu Leu Ala Gly Ala
20 25 30

Ser Ser Gln Arg Ser Val Ala Arg Met Asp Gly Asp Val Ile Ile Gly
35 40 45
Ala Leu Phe Ser Val His His Gln Pro Pro Ala Glu Lys Val Pro Glu
50 55 60
Arg Lys Cys Gly Glu Ile Arg Glu Gln Tyr Gly Ile Gln Arg Val Glu
65 70 75 80
Ala Met Phe His Thr Leu Asp Lys Ile Asn Ala Asp Pro Val Leu Leu
15 85 90 95
Pro Asn Ile Thr Leu Gly Ser Glu Ile Arg Asp Ser Cys Trp His Ser
20 100 105 110
Ser Val Ala Leu Glu Gln Ser Ile Glu Phe Ile Arg Asp Ser Leu Ile
25 115 120 125
Ser Ile Arg Asp Glu Lys Asp Gly Ile Asn Arg Cys Leu Pro Asp Gly
30 130 135 140
Gln Ser Leu Pro Pro Gly Arg Thr Lys Pro Ile Ala Gly Val Ile
35 145 150 155 160
Gly Pro Gly Ser Ser Ser Val Ala Ile Gln Val Gln Asn Leu Leu Gln
40 165 170 175
Leu Phe Asp Ile Pro Gln Ile Ala Tyr Ser Ala Thr Ser Ile Asp Leu
45 180 185 190
Ser Asp Lys Thr Leu Tyr Lys Tyr Phe Leu Arg Val Val Pro Ser Asp
50 195 200 205
Thr Leu Gln Ala Arg Ala Met Leu Asp Ile Val Lys Arg Tyr Asn Trp
5 210 215 220

Thr Tyr Val Ser Ala Val His Thr Glu Gly Asn Tyr Gly Glu Ser Gly
10 225 230 235 240

Met Asp Ala Phe Lys Glu Leu Ala Ala Gln Glu Gly Leu Cys Ile Ala
15 245 250 255

His Ser Asp Lys Ile Tyr Ser Asn Ala Gly Glu Lys Ser Phe Asp Arg
20 260 265 270

Leu Leu Arg Lys Leu Arg Glu Arg Leu Pro Lys Ala Arg Val Val Val
25 275 280 285

Cys Phe Cys Glu Gly Met Thr Val Arg Gly Leu Leu Ser Ala Met Arg
30 290 295 300

Arg Leu Gly Val Val Gly Glu Phe Ser Leu Ile Gly Ser Asp Gly Trp
35 310 315 320

Ala Asp Arg Asp Glu Val Ile Glu Gly Tyr Glu Val Glu Ala Asn Gly
40 325 330 335

Gly Ile Thr Ile Lys Leu Gln Ser Pro Glu Val Arg Ser Phe Asp Asp
45 340 345 350

Tyr Phe Leu Lys Leu Arg Leu Asp Thr Asn Thr Arg Asn Pro Trp Phe
50

55
5  Pro Glu Phe Trp Gln His Arg Phe Gln Cys Arg Leu Pro Gly His Leu
   370 375 380

10 Leu Glu Asn Pro Asn Phe Lys Arg Ile Cys Thr Gly Asn Glu Ser Leu
   385 390 395 400

15 Glu Glu Tyr Val Gln Asp Ser Lys Met Gly Phe Val Ile Asn Ala
   405 410 415

20 Ile Tyr Ala Met Ala His Gly Leu Gln Asn Met His His Ala Leu Cys
   420 425 430

30 Pro Gly His Val Gly Leu Cys Asp Ala Met Lys Pro Ile Asp Gly Ser
   435 440 445

35 Lys Leu Leu Asp Phe Leu Ile Lys Ser Ser Phe Ile Gly Val Ser Gly
   450 455 460

40 Glu Glu Val Trp Phe Asp Glu Lys Gly Asp Ala Pro Gly Arg Tyr Asp
   465 470 475 480

45 Ile Met Asn Leu Gln Tyr Thr Glu Ala Asn Arg Tyr Asp Tyr Val His
   485 490 495

50 Val Gly Thr Trp His Glu Gly Val Leu Asn Ile Asp Asp Tyr Lys Ile
   500 505 510
Gln Met Asn Lys Ser Gly Val Val Arg Ser Val Cys Ser Glu Pro Cys
   515  520  525

5
Leu Lys Gly Gin Ile Lys Val Ile Arg Lys Gly Glu Val Ser Cys Cys
   530  535  540

10
Trp Ile Cys Thr Ala Cys Lys Glu Asn Glu Tyr Val Gln Asp Glu Phe
   545  550  555  560

Thr Cys Lys Ala Cys Asp Leu Gly Trp Trp Pro Asn Ala Asp Leu Thr
   565  570  575

25
Gly Cys Glu Pro Ile Pro Val Arg Tyr Leu Glu Trp Ser Asn Ile Glu
   580  585  590

30
Pro Ile Ile Ala Ile Ala Phe Ser Cys Leu Gly Ile Leu Val Thr Leu
   595  600  605

35
Phe Val Thr Leu Ile Phe Val Leu Tyr Arg Asp Thr Pro Val Val Lys
   610  615  620

40
Ser Ser Ser Arg Glu Leu Cys Tyr Ile Ile Leu Ala Gly Ile Phe Leu
   625  630  635  640

45
Gly Tyr Val Cys Pro Phe Thr Leu Ile Ala Lys Pro Thr Thr Thr Ser
   645  650  655

50
Cys Tyr Leu Gin Arg Leu Leu Val Gly Leu Ser Ser Ala Met Cys Tyr
   660  665  670
Ser Ala Leu Val Thr Lys Thr Asn Arg Ile Ala Arg Ile Leu Ala Gly
\( \text{675} \quad \text{680} \quad \text{685} \)

Ser Lys Lys Lys Ile Cys Thr Arg Lys Pro Arg Phe Met Ser Ala Trp
\( \text{690} \quad \text{695} \quad \text{700} \)

Ala Gln Val Ile Ile Ala Ser Ile Leu Ile Ser Val Gln Leu Thr Leu
\( \text{705} \quad \text{710} \quad \text{715} \quad \text{720} \)

Val Val Thr Leu Ile Ile Met Glu Pro Pro Met Pro Ile Leu Ser Tyr
\( \text{725} \quad \text{730} \quad \text{735} \)

Pro Ser Ile Lys Glu Val Tyr Leu Ile Cys Asn Thr Ser Asn Leu Gly
\( \text{740} \quad \text{745} \quad \text{750} \)

Val Val Ala Pro Leu Gly Tyr Asn Gly Leu Leu Ile Met Ser Cys Thr
\( \text{755} \quad \text{760} \quad \text{765} \)

Tyr Tyr Ala Phe Lys Thr Arg Asn Val Pro Ala Asn Phe Asn Glu Ala
\( \text{770} \quad \text{775} \quad \text{780} \)

Lys Tyr Ile Ala Phe Thr Met Tyr Thr Thr Cys Ile Ile Trp Leu Ala
\( \text{785} \quad \text{790} \quad \text{795} \quad \text{800} \)

Phe Val Pro Ile Tyr Phe Gly Ser Asn Tyr Lys Ile Ile Thr Thr Cys
\( \text{805} \quad \text{810} \quad \text{815} \)

Phe Ala Val Ser Leu Ser Val Thr Val Ala Leu Gly Cys Met Phe Thr
5  Pro Lys Met Tyr Ile Ile Ile Ala Lys Pro Glu Arg Asn Val Arg Ser
   835  840  845

10  Ala Phe Thr Thr Ser Asp Val Val Arg Met His Val Gly Asp Gly Lys
    850  855  860

15  Leu Pro Cys Arg Ser Asn Thr Phe Leu Asn Ile Phe Arg Arg Lys Lys
    865  870  875  880

20  Ala Gly Ala Gly Asn Ala Asn Ser Asn Gly Lys Ser Val Ser Trp Ser
    885  890  895

30  Glu Pro Gly Gly Gly Gln Val Pro Lys Gly Gln His Met Trp His Arg
    900  905  910

35  Leu Ser Val His Val Lys Thr Asn Glu Thr Ala Cys Asn Gln Thr Ala
    915  920  925

40  Val Ile Lys Pro Leu Thr Lys Ser Tyr Gln Gly Ser Gly Lys Ser Leu
    930  935  940

45  Thr Phe Ser Asp Thr Ser Thr Lys Thr Leu Tyr Asn Val Glu Glu Glu
    945  950  955  960

50  Glu Asp Ala Gln Pro Ile Arg Phe Ser Pro Pro Gly Ser Pro Ser Met
    965  970  975
Val Val His Arg Arg Val Pro Ser Ala Ala Thr Thr Pro Pro Leu Pro
980 985 990

Pro His Leu Thr Ala Glu Glu Thr Pro Leu Phe Leu Ala Glu Pro Ala
995 1000 1005

Leu Pro Lys Gly Leu Pro Pro Pro Leu Gln Gln Gln Gln Gln Pro Pro
15 1010 1015 1020

Pro Gln Gln Lys Ser Leu Met Asp Gln Leu Gln Gly Val Val Ser Asn
20 1025 1030 1035 1040

Phe Ser Thr Ala Ile Pro Asp Phe His Ala Val Leu Ala Gly Pro Gly
25 1045 1050 1055

Gly Pro Gly Asn Gly Leu Arg Ser Leu Tyr Pro Pro Pro Pro Pro Pro
30 1060 1065 1070

Gln His Leu Gln Met Leu Pro Leu Gln Leu Ser Thr Phe Gly Glu Glu
35 1075 1080 1085

Leu Val Ser Pro Pro Ala Asp Asp Asp Asp Ser Glu Arg Phe Lys
40 1090 1095 1100

Leu Leu Gln Glu Tyr Val Tyr Glu His Glu Arg Glu Gly Asn Thr Glu
45 1105 1110 1115 1120

Glu Asp Glu Leu Glu Glu Glu Glu Glu Asp Leu Gln Ala Ala Ser Lys
50 1125 1130 1135
Leu Thr Pro Asp Asp Ser Pro Ala Leu Thr Pro Pro Ser Pro Phe Arg
5 1140   1145   1150

Asp Ser Val Ala Ser Gly Ser Ser Val Pro Ser Ser Pro Val Ser Glu
10 1155  1160  1165

Ser Val Leu Cys Thr Pro Pro Asn Val Ser Tyr Ala Ser Val Ile Leu
15 1170  1175  1180

Arg Asp Tyr Lys Gln Ser Ser Ser Thr Leu
20 1185  1190

<210> 6
30 <211> 879
<212> PRT
35 <213> Human adenovirus type 1

<400> 6
40 Met Lys Met Leu Thr Arg Leu Gln Ile Leu Met Leu Ala Leu Phe Ser
1  5  10  15

Lys Gly Phe Leu Leu Ser Leu Gly Asp His Asn Phe Met Arg Arg Glu
20 25   30

Ile Lys Ile Glu Gly Asp Leu Val Leu Gly Gly Leu Phe Pro Ile Asn
55 35   40   45
<table>
<thead>
<tr>
<th>Position</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>Glu Lys Gly Thr Gly Thr Glu Glu Cys Gly Arg Ile Asn Glu Asp Arg</td>
</tr>
<tr>
<td>6-10</td>
<td>Gly Ile Gln Arg Leu Glu Ala Met Leu Phe Ala Ile Asp Glu Ile Asn</td>
</tr>
<tr>
<td>11-15</td>
<td>Lys Asp Asn Tyr Leu Leu Pro Gly Val Lys Leu Gly Val His Ile Leu</td>
</tr>
<tr>
<td>16-20</td>
<td>Asp Thr Cys Ser Arg Asp Thr Tyr Ala Leu Glu Gln Ser Leu Glu Phe</td>
</tr>
<tr>
<td>21-25</td>
<td>Val Arg Ala Ser Leu Thr Lys Val Asp Glu Ala Glu Tyr Met Cys Pro</td>
</tr>
<tr>
<td>26-30</td>
<td>Asp Gly Ser Tyr Ala Ile Gln Glu Asn Ile Pro Leu Leu Ile Ala Gly</td>
</tr>
<tr>
<td>31-35</td>
<td>Val Ile Gly Gly Ser Tyr Ser Ser Val Ser Ile Gln Val Ala Asn Leu</td>
</tr>
<tr>
<td>36-40</td>
<td>Leu Arg Leu Phe Gln Ile Pro Gln Ile Ser Tyr Ala Ser Thr Ser Ala</td>
</tr>
<tr>
<td>41-45</td>
<td>Lys Leu Ser Asp Lys Ser Arg Tyr Asp Tyr Phe Ala Arg Thr Val Pro</td>
</tr>
<tr>
<td>46-50</td>
<td>Pro Asp Phe Tyr Gln Ala Lys Ala Met Ala Glu Ile Leu Arg Phe Phe</td>
</tr>
<tr>
<td>51-55</td>
<td>195 200 205</td>
</tr>
</tbody>
</table>
Asn Trp Thr Tyr Val Ser Thr Val Ala Ser Glu Gly Asp Tyr Gly Glu
5 210 215 220

Thr Gly Ile Glu Ala Phe Glu Gln Glu Ala Arg Leu Arg Asn Ile Cys
10 225 230 235 240

Ile Ala Thr Ala Glu Lys Val Gly Arg Ser Asn Ile Arg Lys Ser Tyr
15 245 250 255

Asp Ser Val Ile Arg Glu Leu Leu Gln Lys Pro Asn Ala Arg Val Val
20 260 265 270

Val Leu Phe Met Arg Ser Asp Asp Ser Arg Glu Leu Ile Ala Ala Ala
25 275 280 285

Asn Arg Val Asn Ala Ser Phe Thr Trp Val Ala Ser Asp Gly Trp Gly
30 290 295 300

Ala Gln Glu Ser Ile Val Lys Gly Ser Glu His Val Ala Tyr Gly Ala
35 305 310 315 320

Ile Thr Leu Glu Leu Ala Ser His Pro Val Arg Gln Phe Asp Arg Tyr
40 325 330 335

Phe Gln Ser Leu Asn Pro Tyr Asn Asn His Arg Asn Pro Trp Phe Arg
45 340 345 350

Asp Phe Trp Glu Gln Lys Phe Gln Cys Ser Leu Gln Asn Lys Arg Asn...
355  360  365

5  His Arg Glu Val Cys Asp Lys His Leu Ala Ile Asp Ser Ser Asn Tyr
370  375  380

10  Glu Gln Glu Ser Lys Ile Met Phe Val Val Asn Ala Val Tyr Ala Met
385  390  395  400

15  Ala His Ala Leu His Lys Met Gln Arg Thr Leu Cys Pro Asn Thr Thr
405  410  415

20  Lys Leu Cys Asp Ala Met Lys Ile Leu Asp Gly Lys Lys Leu Tyr Lys
420  425  430

25  Glu Tyr Leu Leu Lys Ile Asn Phe Thr Ala Pro Phe Asn Pro Asn Lys
435  440  445

30  Gly Ala Asp Ser Ile Val Lys Phe Asp Thr Phe Gly Asp Gly Met Gly
450  455  460

35  Arg Tyr Asn Val Phe Asn Leu Gln Gln Thr Gly Gly Lys Tyr Ser Tyr
465  470  475  480

40  Leu Lys Val Gly His Trp Ala Glu Thr Leu Ser Leu Asp Val Asp Ser
485  490  495

45  Ile His Trp Ser Arg Asn Ser Val Pro Thr Ser Gln Cys Ser Asp Pro
50  500  505  510
<table>
<thead>
<tr>
<th>Residue</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys Ala Pro Asn Glu Met Lys Asn Met Gln Pro Gly Asp Val Cys Cys</td>
<td>515</td>
<td>520</td>
<td>525</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp Ile Cys Ile Pro Cys Glu Pro Tyr Glu Tyr Leu Val Asp Glu Phe</td>
<td>530</td>
<td>535</td>
<td>540</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Cys Met Asp Cys Gly Pro Gly Gln Trp Pro Thr Ala Asp Leu Ser</td>
<td>545</td>
<td>550</td>
<td>555</td>
<td>560</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly Cys Tyr Asn Leu Pro Glu Asp Tyr Ile Lys Trp Glu Asp Ala Trp</td>
<td>565</td>
<td>570</td>
<td>575</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Ile Gly Pro Val Thr Ile Ala Cys Leu Gly Phe Leu Cys Thr Cys</td>
<td>580</td>
<td>585</td>
<td>590</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile Val Ile Thr Val Phe Ile Lys His Asn Asn Thr Pro Leu Val Lys</td>
<td>595</td>
<td>600</td>
<td>605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Ser Gly Arg Glu Leu Cys Tyr Ile Leu Leu Phe Gly Val Ser Leu</td>
<td>610</td>
<td>615</td>
<td>620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Tyr Cys Met Thr Phe Phe Phe Ile Ala Lys Pro Ser Pro Val Ile</td>
<td>625</td>
<td>630</td>
<td>635</td>
<td>640</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys Ala Leu Arg Arg Leu Gly Leu Gly Thr Ser Phe Ala Ile Cys Tyr</td>
<td>645</td>
<td>650</td>
<td>655</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Ala Leu Leu Thr Lys Thr Asn Cys Ile Ala Arg Ile Phe Asp Gly</td>
<td>660</td>
<td>665</td>
<td>670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Val Lys Asn Gly Ala Gln Arg Pro Lys Phe Ile Ser Pro Ser Ser Gln
5 675 680 685

Val Phe Ile Cys Leu Gly Leu Ile Leu Val Gln Ile Val Met Val Ser
10 690 695 700

Val Trp Leu Ile Leu Glu Thr Pro Gly Thr Arg Arg Tyr Thr Leu Pro
15 705 710 715 720

Glu Lys Arg Glu Thr Val Ile Leu Lys Cys Asn Val Lys Asp Ser Ser
20 725 730 735

Met Leu Ile Ser Leu Thr Tyr Asp Val Val Leu Val Ile Leu Cys Thr
25 740 745 750

Val Tyr Ala Phe Lys Thr Arg Lys Cys Pro Glu Asn Phe Asn Glu Ala
30 755 760 765

Lys Phe Ile Gly Phe Thr Met Tyr Thr Thr Cys Ile Ile Trp Leu Ala
35 770 775 780

Phe Leu Pro Ile Phe Tyr Val Thr Ser Ser Asp Tyr Arg Val Gln Thr
40 785 790 795 800

Thr Thr Met Cys Ile Ser Val Ser Leu Ser Gly Phe Val Val Leu Gly
45 805 810 815

Cys Leu Phe Ala Pro Lys Val His Ile Val Leu Phe Gln Pro Gln Lys
Asn Val Val Thr His Arg Leu His Leu Asn Arg Phe Ser Val Ser Gly

Thr Ala Thr Thr Tyr Ser Gln Ser Ser Ala Ser Thr Tyr Val Pro Thr

Val Cys Asn Gly Arg Glu Val Leu Asp Ser Thr Thr Ser Ser Leu

<210> 7
<211> 1467
<212> DNA
<213> Human adenovirus type 1

gtactacaac ctgtgtgacc gctcttgagg gcgctggtgg gtcggtgggtc gctcttgagg cctggtgtgg 60
gccgaggtt gcggcgatg tttggtggcc caaaaggttgg tggcggagcg tggcagttggt 120
gcaggacacc aagaaacggag ccttgctggag cacccagggta ttcttccttc tgtggaacc 180
tggccctcct ttctggttcct tgtttgcttg tgtgtggtgaag ccgcaccttc tcccacagt 240
cctcgtggcg cctcctcttt ggggtctgtg tggcgcctctg tgtctcttggt ctggcggctc 300
acgctcttttc cctcaacgcg tggcggcggag aagaaacggag ggcggcgggg tgtttgatgt 360
tccctgtggc tgtctgtgtgc accctggttag aggtcatcat caatacagag tgtggtgtca 420
tccctgtgtt tggggctagt gggtgcggttgg cctgttgccgg gcatacgccg ccggtggtgg 480
cctggccctc cccctgtgcc atcgccaaac tggactttgtg catggcactc atctcagttca 540
tgcctgtcgct gcggctgctg gctcttgagg ccctggtggcc gctgttggtgc gcgtacaagc 600
gctggcggtaa gnatggggtct tttgtgtcctc tcaacaagc cactcccttg gccataggg 660
tggttggttt ctgcattgat acctacggca acaagcagca caacagtcct acctgggtatg 720
acccccacgct ggccatcgcc ctgcgcgcgc aatgcctggtc cttcgtcttc ttctacgtca 780
5
ttcctcagggt tctccatggtg accaagtctca gcggcagagc aagctaccag ggagcatgtg 840
acccccccgg gggccgttggg ctatgagacc atcctgaaag agcagaaggg tcagacgcatg 900
ttcgtggaga acaagccctt tctccatggtg gcagcggggtg cagctaaagag gcgggttgtca 960
ccataccagc ggtacaaatgg gcagcgtgtgtg accaatgtgtg accaggccccac tgaagatggcc 1020
cctgatgcca aagttccgtgc ggaagagctt tagcatactca tcctccacgc gcggacccgc 1080
15
aacagccagg tggtagggcag tgcacactgca cctcgcggct gaagacatgt actcggcccc 1140
gagccaccag ggacccaccc gcggagaaga gcggcagaacc tcctcaggtct ttwagaaacc 1200
cctacgttgtt ggagctgagtc agcggtggcc agggagaggg ggtccggattt yggggaggg 1260
cctcgagac ccggccccgg gcggagagat tytcgcaggyt tctcctcccc ctggcagggc 1320
cagcaacatg tgccccagat gtggagggcc cttctctctt gccaatgttt gggtgggttgt 1380
25
caatgggtct cccaccccact cctcagttgtg tgtggagtcc ggagagccac cccagccctc 1440
tgcaggagtc acctcggccgg tcacact 1467

1
35

1
CLAIMS:

1. A substantially purified MGlur-like receptor protein MGRcm comprising a sequence having at least 80% sequence identity to SEQ ID NO:2 or to SEQ ID NO:4.

2. The MGRcm protein of claim 1, comprising a sequence having at least 90% sequence identity to SEQ ID NO:2 or SEQ ID NO:4.

3. The MGRcm protein of Claim 1, comprising the sequence SEQ ID NO:2 or SEQ ID NO:4.

4. An isolated polynucleotide comprising
   (a) a sequence encoding the MGRcm protein of Claim 1, or
   (b) a sequence complementary to the sequence of (a).

5. The polynucleotide of Claim 4, comprising a sequence having at least 80% identity to SEQ ID NO:1 or SEQ ID NO:3.

6. The polynucleotide of Claim 5 comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:3.

7. An expression vector containing the polynucleotide of Claim 4, and regulatory elements effective for expression of the polynucleotide in a suitable host.

8. An expression vector containing the polynucleotide of Claim 6, and regulatory elements effective for expression of the polynucleotide in a suitable host.

9. A host cell transfected with the expression vector of Claim 7.

10. A host cell cell transfected with the expression vector of Claim 8.

11. A purified antibody which specifically binds to the protein of Claim 1.

12. A method for detecting a polynucleotide which encodes an MGRcm protein in a biological sample, comprising the steps of:

   (a) hybridizing to nucleic acid material of said biological sample a polynucleotide fragment derived from the sequence identified as SEQ ID NO:1 or SEQ ID NO:3, said fragment having a length of at least 15 and 30 nucleotides, thereby forming a hybridization complex; and
(b) detecting said hybridization complex;

wherein the presence of said hybridization complex correlates with the presence of a polynucleotide encoding the MGRcm protein in said biological sample.

13. A method of identifying a modulator of MGRcm, the method comprising:
(a) contacting a test compound with a protein according to Claim 1;
(b) measuring the effect of the test compound on the physiological activity of MGRcm,
(c) selecting the test compound as a candidate compound if its effect on the physiological activity of MGRcm protein above a selected threshold level.

14. A method according to Claim 13, wherein said physiological activity is signal transduction.

15. A method according to Claim 13, wherein the effect is increase in physiological activity and the modulator is an agonist.

16. A method according to Claim 13, wherein the effect is decrease in physiological activity and the modulator is an antagonist.

17. An agonist obtained by the method of Claim 15.

18. An antagonist obtained by the method of Claim 16.

19. A method of identifying a compound which inhibits binding of an MGRcm agonist to MGRcm, the method comprising:
(a) incubating together a test compound, MGRcm, and the MGRcm agonist, under conditions which allow the MGRcm agonist bind to MGRcm in the absence of said test compound,
(b) measuring the extent of binding of the MGRcm agonist to MGRcm in the presence of the test compound, and
(c) identifying the test compound as a candidate compound if the extent of binding of the agonist to MGRcm in the presence of the test compound is less than the extent of binding of the agonist to MGRcm in the absence of the test compound.
20. A candidate compound identified by the method of Claim 19.

21. A method for detecting MGRcm in a biological sample, comprising the steps of:

   (a) contacting with said biological sample the antibody of claim 9, thereby forming an antibody-antigen complex; and

   (b) detecting said antibody-antigen complex;

   wherein the presence of said antibody-antigen complex correlates with the presence of MGRcm in said biological sample.
Fig. 1A
Fig. 1B
Fig. 2A
Fig. 2B
Fig. 3A
Fig. 3B
Fig. 4
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

<table>
<thead>
<tr>
<th>IPC</th>
<th>C12N15/12</th>
<th>C07K14/05</th>
<th>C12Q1/68</th>
<th>G01N33/53</th>
<th>G01N33/68</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C07K16/28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

<table>
<thead>
<tr>
<th>IPC</th>
<th>C12N</th>
<th>C07K</th>
<th>C12Q</th>
<th>G01N</th>
</tr>
</thead>
</table>

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P,X</td>
<td>WO 99 06550 A (GENSET (FR); DUMAS MILNE EDWARDS JEAN-BAPTISTE; DUCLERT A.; LACROIX B.) 11 February 1999 (1999-02-11)</td>
<td>1-21</td>
</tr>
<tr>
<td></td>
<td>page 6, line 23 - page 12, line 29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 78, line 29 - page 79, line 12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 85, line 26 - page 86, line 9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 91, line 23-31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 108, line 11-22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 130 - page 134; claims</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seq.ID:220 is 99% identical to Seq.ID:1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>nt.1436-1641</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 200 - page 201</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seq.ID:498 is 100% identical to Seq.ID:2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>aa.307-365</td>
<td></td>
</tr>
<tr>
<td></td>
<td>page 470 - page 471</td>
<td></td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents:
  *A* document defining the general state of the art which is not considered to be of particular relevance
  *E* earlier document but published on or after the international filing date
  *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  *O* document referring to an oral disclosure, use, exhibition or other means
  *P* document published prior to the international filing date but later than the priority date claimed

*"T"* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

*"X"* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

*"Y"* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

*"A"* document member of the same patent family

Date of the actual completion of the international search

26 August 1999

Date of mailing of the international search report

2 09 99

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentgaat
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
Fax: (+31-70) 340-3016

Authorized officer

Macchia, G
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>
| E        | WO 99 31117 A (HUMAN GENOME SCIENCES INC (US) MOORE RUBEN CARTER SHI ROSEN)  
Gene No.79; Seq.ID:89, 124, 203, 238  
Seq.ID:203 is 100% identical to Seq.ID:2  
aa.46-429  
page 182  
page 188, line 27-30  
page 201, line 30 - page 206, line 26  
page 212, line 34 - page 214, line 7  
page 266 - page 269; claims | 1-21 |
| X        | EP 0 679 716 A (MATSUBARA KENICHI (JP); OKUBO KOUSAKU (JP))  
Seq.ID:6645 is 95% identical to Seq.ID:1  
nt.2346-2404 | 1-12,21 |
| X        | KNÖPFEL T. AND GASPARINI F.: "Metabotropic Glutamate Receptors:  
potential drug targets"  
JOURNAL OF MEDICINAL CHEMISTRY,  
vol. 1, 1 January 1996 (1996-01-01), pages  
103-108, XP000650306  
ISSN: 1359-6446 | 17,18,20 |
| A        | the whole document | 13-16,19 |
| X        | KNÖPFEL T. ET AL.: "Metabotropic  
Glutamate Receptors: novel targets for  
drug development"  
NEUROPHARMACOLOGY,  
vol. 34, no. 1, 1995, pages  
1-26, XP002113327 | 17,18,20 |
| A        | PIN J.-P. AND DUVOISIN R.: "Neurotransmitter receptors I. The  
metabotropic glutamate receptors:  
structure and functions"  
JOURNAL OF MEDICINAL CHEMISTRY,  
vol. 1, 1 January 1996 (1996-01-01), pages  
103-108, XP000650306  
ISSN: 1359-6446 | 13-16,19 |
| A        | WO 96 06167 A (CIBA GEIGY AG (CH); FLOR  
P.J.; KUHN R.; LINDAUER K.; PÜTTLER I.;  
KNÖPFEL T) 29 February 1996 (1996-02-29) | ----- |
**INTERNATIONAL SEARCH REPORT**

**Box I**  Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. □ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. X Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
   see FURTHER INFORMATION sheet PCT/ISA/210

3. □ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II**  Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. □ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. □ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. □ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. □ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

□ The additional search fees were accompanied by the applicant's protest.

□ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)
Continuation of Box I.2

Present claims 17, 18, 20 relate to an extremely large number of possible compounds. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the compounds claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to peptides derived from the disclosed polypeptides, antibodies and antisense molecules.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>WO 9906550 A</td>
<td>11-02-1999</td>
<td>AU 8555198 A</td>
<td>22-02-1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 9931116 A</td>
<td>24-06-1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 6552198 A</td>
<td>29-09-1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 9840483 A</td>
<td>17-09-1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2153480 A</td>
<td>01-06-1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 9514772 A</td>
<td>01-06-1995</td>
</tr>
<tr>
<td>WO 9606167 A</td>
<td>29-02-1996</td>
<td>AU 3110095 A</td>
<td>14-03-1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2196997 A</td>
<td>29-02-1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0776364 A</td>
<td>04-06-1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FI 970633 A</td>
<td>14-02-1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HU 76969 A</td>
<td>28-01-1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO 970740 A</td>
<td>14-04-1997</td>
</tr>
</tbody>
</table>